

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
17. Juni 2004 (17.06.2004)

PCT

(10) Internationale Veröffentlichungsnummer
WO 2004/051502 A2

(51) Internationale Patentklassifikation⁷: **G06F 17/21**

(72) Erfinder; und

(21) Internationales Aktenzeichen: PCT/EP2003/013511

(75) Erfinder/Anmelder (nur für US): HEUER, Jörg
[DE/DE]; Fischbachauerstr. 8, 81539 München (DE).
HUTTER, Andreas [DE/DE]; Sturmiusweg 42, 81673
München (DE).

(22) Internationales Anmeldedatum:
1. Dezember 2003 (01.12.2003)

(74) Gemeinsamer Vertreter: SIEMENS AKTIENGESELLSCHAFT; Postfach 22 16 34, 80506 München (DE).

(25) Einreichungssprache: Deutsch

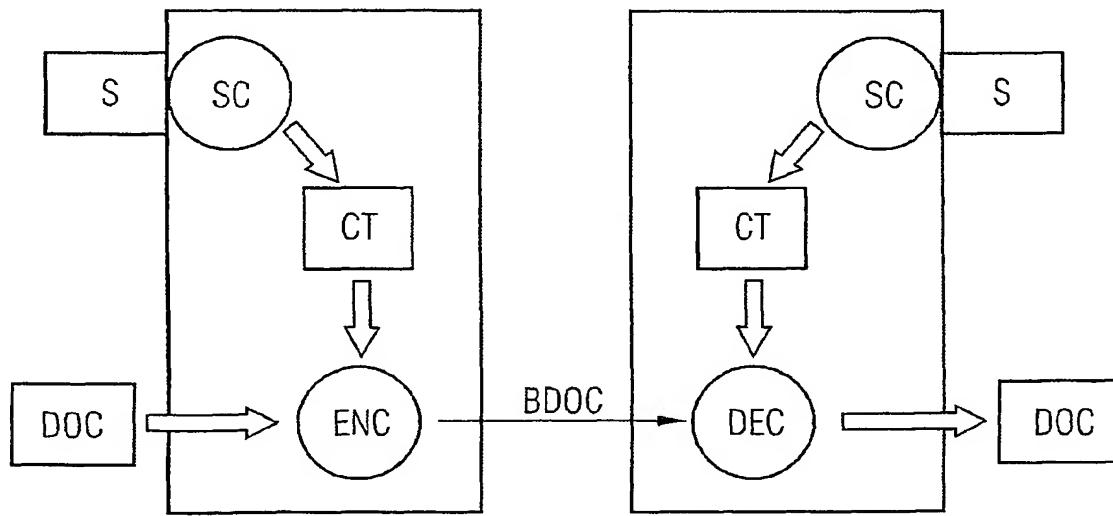
(81) Bestimmungsstaaten (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GI, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(26) Veröffentlichungssprache: Deutsch

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD FOR ENCODING AN XML-BASED DOCUMENT

(54) Bezeichnung: VERFAHREN ZUR CODIERUNG EINES XML-BASIERTEN DOKUMENTS



(57) Abstract: The invention relates to a method for encoding an XML-based document (DOC), the contents of said document corresponding to an XML-schema voice definition. According to said method, an encoded binary representation (BDOC) of the document is produced by associating the contents of the document with binary structural codes (SBC) using encoding tables (CT), textual contents of a "complex type" data type being associated with the "mixed" content model structural codes (SBC).

WO 2004/051502 A2

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren zur Codierung eines XML-basierten Dokuments (DOC), das Inhalte gemäß einer XML-Schema-Sprachdefinition enthält, bei dem eine codierte Binärdarstellung (BDOC) des Dokuments erzeugt wird, indem den Inhalten des Dokuments binäre Struktur-Codes (SBC) über Codiertabellen (CT) zugeordnet werden, wobei textuellen Inhalten eines Datentyps "Complex Type" mit dem Inhaltsmodell "mixed" Struktur-Codes (SBC) zugeordnet werden.



(84) **Bestimmungsstaaten (regional):** ARIPO-Patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), curasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI-Patent (BF, BJ, CE, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

— ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts

Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

JC17 Rec'd PCT/PTO 03 JUN 2005

1

Beschreibung

Verfahren zur Codierung eines XML-basierten Dokuments

- 5 Die Erfindung betrifft ein Verfahren zur Codierung eines XML-basierten Dokuments, das Inhalte gemäß einer XML-Schema-Sprachdefinition enthält, ein entsprechendes Decodierverfahren sowie entsprechende Codier- und Decodierzvorrichtungen.
- 10 XML (= extensible markup language) ist eine Sprache, mit der eine strukturierte Beschreibung der Inhalte eines Dokuments mittels XML-Schema-Sprachdefinitionen ermöglicht wird. Eine genauere Beschreibung des XML-Schemas sowie der darin verwendeten Strukturen, Datentypen und Inhaltsmodelle findet sich
15 in den Referenzen [1], [2] und [3].

Aus dem Stand der Technik sind Verfahren zur Codierung von XML-basierten Dokumenten bekannt, bei denen das Dokument in eine codierte Binärdarstellung umgewandelt wird. Beispielsweise werden in dem Dokument [4], das im Rahmen der Entwicklung eines MPEG-7-Codierstandards entstanden ist, Verfahren zur Codierung und Decodierung von XML-basierten Dokumenten beschrieben.

- 25 Die aus dem Stand der Technik bekannten Verfahren zur Erzeugung einer Binärdarstellung von XML-basierten Dokumenten weisen Nachteile bei der Codierung von "Complex Type"-Datentypen mit dem Inhaltsmodell "mixed" auf, da diese Datentypen neben Elementen textuelle Inhalte enthalten können, die jedoch nur
30 durch die Decodierung des gesamten Datenstroms rekonstruiert werden können. Eine nähere Beschreibung des Datentyps "Complex Type" sowie des Inhaltsmodells "mixed" findet sich in Dokument [1].
- 35 Aufgabe der Erfindung ist es deshalb, ein Verfahren zur Codierung von XML-basierten Dokumenten zu schaffen, welches einen einfacheren Zugriff auf codierte textuelle Inhalte des

Datentyps "Complex Type" mit Inhaltsmodell "mixed" ermöglicht.

5 Diese Aufgabe wird durch die unabhängigen Patentansprüche ge-
löst. Weiterbildungen der Erfindung sind in den abhängigen
Ansprüchen definiert.

Bei dem erfindungsgemäßen Codierverfahren wird eine codierte
10 Binärdarstellung eines XML-basierten Dokuments erzeugt, indem
den Inhalten des Dokuments binäre Struktur-Codes über Codier-
tabellen zugeordnet werden, wobei textuellen Inhalten eines
Datentyps "Complex Type" mit dem Inhaltsmodell "mixed" Struk-
15 tur-Codes zugeordnet werden. Bei den Struktur-Codes handelt
es sich um die im Abschnitt 7.6.1 des Dokuments [4] definier-
ten Schema-Branch-Codes SBC. Durch die in [4] beschriebene
Zuordnung von Struktur-Codes zu Inhalten des Dokuments kann
die Lage dieser Inhalte in der Struktur der XML-Dokumente
signalisiert oder adressiert werden.

20 Die Erfindung besteht im Wesentlichen darin, dass der tex-
tuelle Inhalt eines Typs „Complex Type“ mit Inhaltsmodell
„mixed“ bei der Codezuweisung wie eine Elementdeklaration in
der Typendefinition betrachtet wird. Dementsprechend wird für
25 die Codierung neben den deklarierten Elementen in einer Typ-
definition zusätzlich auch dem textuellen Inhalt ein festge-
legter Struktur-Code zugewiesen, wenn für den Typ ein In-
haltsmodell mixed definiert ist. Hierdurch werden textuelle
Inhalte im codierten Datenstrom adressiert, so dass auf diese
30 Inhalte zugegriffen werden kann, ohne dass der gesamte Daten-
strom decodiert werden muss.

In einer bevorzugten Ausführungsform des erfindungsgemäßen
Codierverfahrens erfolgt die Zuordnung der Struktur-Codes zu
den textuellen Inhalten eines Datentyps "Complex Type" mit
35 Inhaltsmodellen "mixed" ausschließlich über OperandTBC-
Codiertabellen. Diese Codiertabellen legen die Codes der sog.
OperandTBCs, d.h. der sog. TBCs (TBC = Tree Branch Code) der

sog. Operand Nodes fest. Eine genaue Beschreibung und die Definitionen der OperandTBCs und Operand Nodes findet sich in den Abschnitten 7.6.1 sowie 7.6.5.2 des Dokuments [4].

- 5 In einer besonders bevorzugten Ausführungsform werden den textuellen Inhalten eines Datentyps „Complex Type“ mit dem Inhaltsmodell „mixed“ ferner „Position Codes“ zugeordnet. Es handelt sich hierbei um die im Abschnitt 7.6.5.5 des Dokuments [4] näher beschriebenen Position Codes. Da in einem Datentyp „Complex Type“ mit dem Inhaltsmodell „mixed“ mehrere textuelle Inhalte enthalten sein können, wird hierdurch die Information übertragen, an welcher Stelle sich die textuellen Inhalte innerhalb des Datentyps befinden.
- 10 15 In einer besonders bevorzugten Ausführungsform werden bei der Zuordnung der „Position Codes“ „Single Element Position Codes“ und/oder „Multiple Element Position Codes“ verwendet. Diese Position Codes sind in der Druckschrift [4], Abschnitt 7.6.5.5 näher beschrieben. Single Element Position Codes werden insbesondere dann verwendet, wenn keine „Model Group“ in der Typendefinition des „Complex Type“ in der XML-Schema-Definition öfters als einmal auftreten kann. Eine Definition der „Model Group“ findet sich in Dokument [2]. Der Single Element Position Code bestimmt hierbei die Position eines Inhalts bezüglich eines jeweiligen Partikels in einer Instanziierung eines Datentyps. Eine Definition für Partikel befindet sich ebenfalls im Dokument [2]. Der Single Element Position Code wird unter der Annahme codiert, dass der textuelle Inhalt maximal MPA+1-mal deklariert ist, wobei MPA die Anzahl aller in diesem Datentypen möglichen Partikel-Instanziierungen bezeichnet. Ein Multiple Element Position Code wird dann verwendet, wenn in der Definition des „Complex Type“ „Model Groups“ in der XML-Schema-Definition öfters als einmal auftreten können. Der Multiple Element Position Code wird unter der Annahme codiert, dass insgesamt $2 \times MPA + 1$ Positionen adressiert werden können, wobei dieser Code die Posi-
- 20 25 30 35

tion des Inhalts bezüglich aller Partikel in einer Instanziierung eines Datentyps wiedergibt.

In einer weiteren bevorzugten Ausführungsform werden die Position Codes mit Codes variabler Länge codiert, insbesondere mit dem Code vluimsbf5, der in Dokument [4], Abschnitt 4.3 beschrieben ist.

Neben dem oben beschriebenen Codierverfahren umfasst die Erfindung ferner ein Decodierverfahren, mit dem eine gemäß dem oben beschriebenen Codierverfahren codierte Binärdarstellung eines XML-basierten Dokuments decodiert wird. Bei diesem Decodierverfahren werden binäre Repräsentationen von textuellen Inhalten eines Datentyps „Complex Type“ mit dem Inhaltsmodell „mixed“, denen bei der Codierung Struktur-Codes (SBC) zugeordnet wurden, in die den Struktur-Codes (SBC) zugeordneten textuellen Inhalte des XML-basierten Dokuments umgewandelt.

In Analogie zum Codierverfahren erfolgt in einer bevorzugten Ausführungsform die Zuordnung durch Struktur-Codes (SBC) über OperandTBC-Codiertabellen.

In einer bevorzugten Ausführungsform werden ferner binäre Repräsentationen von textuellen Inhalten eines Datentyps „Complex Type“ mit dem Inhaltsmodell „mixed“, adressiert mit „Position Codes“, in textuelle Inhalte an der zugeordneten Position umgewandelt. Die „Position Codes“ können hierbei wiederum „Single Element Position Codes“ und/oder „Multiple Element Position Codes“ umfassen. Es handelt sich bei diesen Position Codes um die gleichen Position Codes, wie sie in Bezug auf das Codierverfahren definiert sind. In Analogie zum Codierverfahren können die "Position Codes" ferner mit Codes variabler Länge codiert sein, wobei diese Codes bei der Umwandlung der Position Codes in textuelle Inhalte decodiert werden. Vorzugsweise sind die Position Codes mit dem Code vluimsbf5 codiert.

Neben den oben beschriebenen Codier- bzw. Decodierverfahren umfasst die Erfindung ferner ein Codier- und Decodierverfahren, welches das erfindungsgemäße Codierverfahren und das erfindungsgemäße Decodierverfahren umfasst.

5

Darüber hinaus betrifft die Erfindung eine Vorrichtung zur Codierung von XML-basierten Dokumenten, mit denen das erfindungsgemäße Codierverfahren durchführbar ist, wobei die Vorrichtung ein Speichermittel umfasst, in dem wenigstens eine

10 Zuordnung eines textuellen Inhalts eines Datentyps "Complex Type" mit dem Inhaltsmodell "mixed" zu einem Struktur-Code gespeichert ist. Die Erfindung betrifft analog eine Vorrichtung zur Decodierung einer codierten Binärdarstellung eines XML-basierten Dokuments, wobei die Vorrichtung derart eingerichtet ist, dass das erfindungsgemäße Decodierverfahren

15 durchführbar ist. Die Vorrichtung umfasst ein Speichermittel, in dem wenigstens eine Zuordnung eines Struktur-Codes zu einem textuellen Inhalt eines Datentyps "Complex Type" mit dem Inhaltsmodell "mixed" gespeichert ist.

20

Darüber hinaus betrifft die Erfindung eine Vorrichtung zur Codierung und Decodierung eines XML-basierten Dokuments, umfassend die oben beschriebene erfindungsgemäße Codervorrichtung und die oben beschriebene erfindungsgemäße Decodervorrichtung.

25

Ausführungsbeispiele der Erfindung werden nachfolgend anhand der beigefügten Zeichnungen erläutert.

30 Es zeigen:

Figur 1 eine Prinzipdarstellung eines erfindungsgemäßen Codier- und Decodiersystems mit Encoder und Decoder;

35 Figur 2 eine Darstellung einer XML-Schema-Definition, in der u.a. ein Datentyp „Complex Type“ mit Inhaltsmodell „mixed“ definiert ist;

Figur 3 eine Darstellung eines XML-Dokuments, in dem ein in der XML-Schema-Definition der Fig. 2 deklariertes Element „MixedElement“ instanziert wird;

5

Figur 4 eine graphische Darstellung des Aufbaus des in dem XML-Dokument der Fig. 3 instanziierten Elements „MixedElement“;

10 Figur 5 eine Darstellung zur Erläuterung der Zuweisung von Struktur-Codes bei Datentypen „Complex Type“ mit Inhaltsmodell „mixed“; und

15 Figur 6 eine Darstellung zur Erläuterung der Zuweisung von „Position Codes“ bei Datentypen „Complex Type“ mit Inhaltsmodell „mixed“.

In Figur 1 ist beispielhaft ein Codier- und Decodiersystem mit einem Encoder ENC und einem Decoder DEC dargestellt, mit 20 denen XML-Dokumente DOC codiert bzw. decodiert werden. Sowohl der Encoder als auch der Decoder verfügen beide über ein so- genanntes XML-Schema S, in dem die zur Kommunikation genutzten Elemente und Typen des XML-Dokuments deklariert und definiert sind. Aus dem Schema S werden über entsprechende Sche- 25 ma-Compilationen SC im Encoder und Decoder Code-Tabellen CT erzeugt. Wenn das XML-Dokument DOC codiert wird, werden den Inhalten des XML-Dokuments über die Code-Tabellen binäre Co- des zugeordnet. Hierdurch wird eine Binärdarstellung BDOC des Dokuments DOC erzeugt, die mithilfe der Code-Tabelle CT im 30 Decoder wieder decodiert werden kann.

Das erfindungsgemäße Verfahren zeichnet sich dadurch aus, dass textuellen Inhalten eines Datentyps „Complex Type“ mit dem Inhaltsmodell „mixed“ binäre Struktur-Codes zugeordnet 35 werden. Dies ermöglicht ein Herausfiltern der textuellen Da- ten aus der Binärdarstellung BDOC, ohne dass die gesamte Bi- närdarstellung BDOC decodiert werden muss.

In Figur 2 ist beispielhaft ein Schema S dargestellt, wobei in diesem Schema in den Zeilen 4 bis 10 ein Element mit dem Namen „Example“ deklariert ist, das wiederum ein Element vom 5 Namen „MixedElement“ vom Typ „MixedType“ enthält. In den Zeilen 12 bis 17 ist der Typ „MixedType“ definiert. Es handelt sich hierbei um einen „Complex Type“ Datentyp mit dem Inhaltsmodell „mixed“, was insbesondere der Zeile 12 entnommen werden kann. Der Typ „MixedType“ enthält zwei Elemente mit 10 den Namen „firstElement“ und „secondElement“, die beide vom Typ „string“ sind.

In Figur 3 ist eine Instanziierung des Elements "MixedElement" in einem XML-Dokument dargestellt. Da das Inhaltsmodell 15 "mixed" textuelle Inhalte in Form von strings enthalten kann, können vor, nach oder zwischen den ersten und zweiten Elementen „firstElement“ und „secondElement“ textuelle Inhalte auftreten. In dem Beispiel der Figur 3 treten insgesamt drei textuelle Inhalte auf.

20 In Figur 4 ist die Struktur des Elements „MixedElement“, das in Figur 3 instanziert ist, nochmals anschaulich als Baumstruktur dargestellt. Von dem obersten MixedElement/mixedType-Knoten hängen in einer ersten Hierarchieebene fünf weitere Knoten ab, welche sowohl die textuellen Inhalte als auch die Elemente „firstElement“ bzw. „secondElement“ enthalten. Die Elemente „firstElement“ und „secondElement“ enthalten in einer zweiten Hierarchieebene ferner die entsprechenden Inhalte „Content of firstElement“ bzw. „Content 25 30 of secondElement“.

Ein beliebiges Dokument basierend auf der XML-Sprache kann durch eine sog. Baumstruktur dargestellt werden, wobei die Inhalte des XML-Dokuments Knoten in der Baumstruktur bilden 35 und sog. Context-Pfade zu diesen Knoten führen. Den Knoten der Baumstruktur werden bei der Codierung binäre Struktur-Codes zugewiesen.

Nach dem Stand der Technik werden für den in Figur 4 gezeig-
ten Elementknoten „MixedElement“ ein Struktur-Code für den
Vaterknoten sowie für die Elemente „firstElement“ und
5 „secondElement“ zugewiesen. Der Vaterknoten ist hierbei der
Knoten, der in der nächsthöheren Hierarchieebene mit dem Kno-
ten des Elements „MixedElement“ verbunden ist. Im Unterschied
hierzu wird gemäß dem erfindungsgemäßen Verfahren nicht nur
ein Struktur-Code für den Vaterknoten und die Elemente
10 „firstElement und „secondElement“ vergeben, sondern es wird
ferner ein Struktur-Code für den textuellen Inhalt zugewie-
sen. Dies ist in Figur 5 verdeutlicht, wobei der Code 00 dem
Vaterknoten zugewiesen wird, der Code 01 dem textuellen In-
halt zugewiesen wird und die Codes 10 bzw. 11 dem „firstEle-
15 ment“ bzw. dem „secondElement“ zugewiesen werden.

In dem erfindungsgemäßen Verfahren ist es ferner möglich, den
einzelnen textuellen Inhalten auch noch „Position Codes“ zu-
zuweisen, wie in Figur 6 dargestellt ist. Da insgesamt an
20 drei Positionen textuelle Inhalte auftreten können, werden
hierzu drei „Position Codes“ benötigt, wobei gemäß Figur 6
die Codes 00, 01 und 10 verwendet werden.

Literaturverzeichnis:

[1] <http://www.w3.org/TR/2001/REC-xmlschema-0-20010502/>

5 [2] <http://www.w3.org/TR/2001/REC-xmlschema-1-20010502/>

[3] <http://www.w3.org/TR/2001/REC-xmlschema-2-20010502/>

10 [4] ISO/IEC FDIS 15938-1 "Information Technology - Multimedia Content Description Interface -Part 1: Systems", Geneva 2002

Patentansprüche

1. Verfahren zur Codierung eines XML-basierten Dokuments (DOC), das Inhalte gemäß einer XML-Schema-Sprachdefinition enthält, bei dem:

5 eine codierte Binärdarstellung (BDOC) des Dokuments erzeugt wird, indem den Inhalten des Dokuments binäre Struktur-Codes (SBC) über Codiertabellen (CT) zugeordnet werden, wobei textuellen Inhalten eines Datentyps „Complex Type“ mit dem Inhaltsmodell „mixed“ Struktur-Codes (SBC) 10 zugeordnet werden.

2. Verfahren nach Anspruch 1, bei dem die Zuordnung der Struktur-Codes (SBC) zu den textuellen Inhalten eines Datentyps „Complex Type“ mit Inhaltsmodell „mixed“ ausschließlich über OperandTBC-Codiertabellen erfolgt.

15 3. Verfahren nach Anspruch 1 oder 2, bei dem den textuellen Inhalten eines Datentyps „Complex Type“ mit dem Inhaltsmodell „mixed“ ferner „Position Codes“ zugeordnet werden.

20 4. Verfahren nach Anspruch 3, bei dem bei der Zuordnung der „Position Codes“ „Single Element Position Codes“ (SPC) und/oder „Multiple Element Position Codes“ (MPC) verwendet werden.

25 5. Verfahren nach Anspruch 3 oder 4, bei dem die „Position Codes“ mit Codes variabler Länge codiert werden.

30 6. Verfahren nach Anspruch 5, bei dem die „Position Codes“ mit dem Code vluimsbf5 codiert werden.

35 7. Verfahren zur Decodierung einer mit einem Verfahren nach einem der vorhergehenden Ansprüche codierten Binärdarstellung eines XML-basierten Dokuments (DOC), bei dem binäre Repräsentationen von textuellen Inhalten eines Datentyps „Complex Type“ mit dem Inhaltsmodell „mixed“, denen bei

11

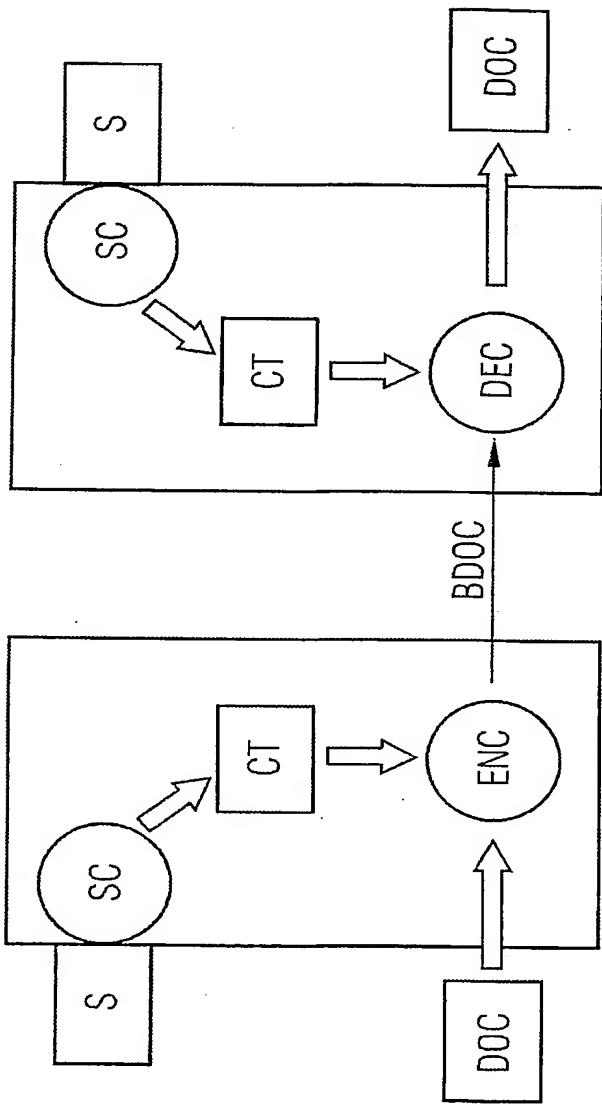
der Codierung Struktur-Codes (SBC) zugeordnet wurden, in die den Struktur-Codes (SBC) zugeordneten textuellen Inhalten des XML-basierten Dokuments umgewandelt werden.

- 5 8. Verfahren nach Anspruch 7, bei dem die Zuordnung durch Struktur-Codes (SBC) über OperandTBC-Codiertabellen erfolgt.
- 10 9. Verfahren nach Anspruch 7 oder 8 zur Decodierung einer mit einem Verfahren nach einem der Ansprüche 3 bis 6 codierten Binärdarstellung eines XML-basierten Dokuments (DOC), bei dem ferner binäre Repräsentationen von textuellen Inhalten eines Datentyps „Complex Type“ mit dem Inhaltsmodell „mixed“, adressiert mit „Position Codes“, in textuelle Inhalte an der zugeordneten Position umgewandelt werden.
- 15 10. Verfahren nach Anspruch 9, bei dem die „Position Codes“ „Single Element Position Codes“ (SPC) und/oder „Multiple Element Position Codes“ (MPC) umfassen.
- 20 11. Verfahren nach Anspruch 9 oder 10, bei dem die „Position Codes“ mit Codes variabler Länge codiert sind.
- 25 12. Verfahren nach Anspruch 11, bei dem die „Position Codes“ mit dem Code vluimsbf5 codiert sind.
- 30 13. Verfahren zur Codierung und Decodierung von XML-basierten Dokumenten, umfassend ein Verfahren nach einem der Ansprüche 1 bis 6 und ein Verfahren nach einem der Ansprüche 7 bis 12.
- 35 14. Vorrichtung zur Codierung von XML-basierten Dokumenten gemäß einem Verfahren nach einem der Ansprüche 1 bis 6, umfassend ein Speichermedium, in dem wenigstens eine Zuordnung eines textuellen Inhalts eines Datentyps „Complex Type“ mit dem Inhaltsmodell „mixed“ zu einem Struktur-Code (SBC) gespeichert ist.

15. Vorrichtung zur Decodierung einer codierten Binärdarstellung eines XML-basierten Dokuments gemäß einem Verfahren nach einem der Ansprüche 7 bis 12, umfassend ein Speichermittel, in dem wenigstens eine Zuordnung eines Struktur-Codes (SBC) zu einem textuellen Inhalt eines Datentyps „Complex Type“ mit dem Inhaltsmodell „mixed“ gespeichert ist.
5
- 10 16. Vorrichtung zur Codierung und Decodierung eines XML-basierten Dokuments (DOC), umfassend die Vorrichtung nach Anspruch 14 und die Vorrichtung nach Anspruch 15.

1/4

FIG 1



THIS PAGE BLANK (USPTO)

2/4

FIG 2

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementForm-
Default="qualified" attributeFormDefault="unqualified">
<xs:element name="Example">
<xs:complexType>
<xs:sequence>
<xs:element name="MixedElement" type="MixedType"/>
</xs:sequence>
</xs:complexType>
</xs:element>

<xs:complexType name="MixedType" mixed="true">
<xs:sequence>
<xs:element name="firstElement" type="xs:string"/>
<xs:element name="secondElement" type="xs:string"/>
</xs:sequence>
</xs:complexType>
</xs:schema>
```

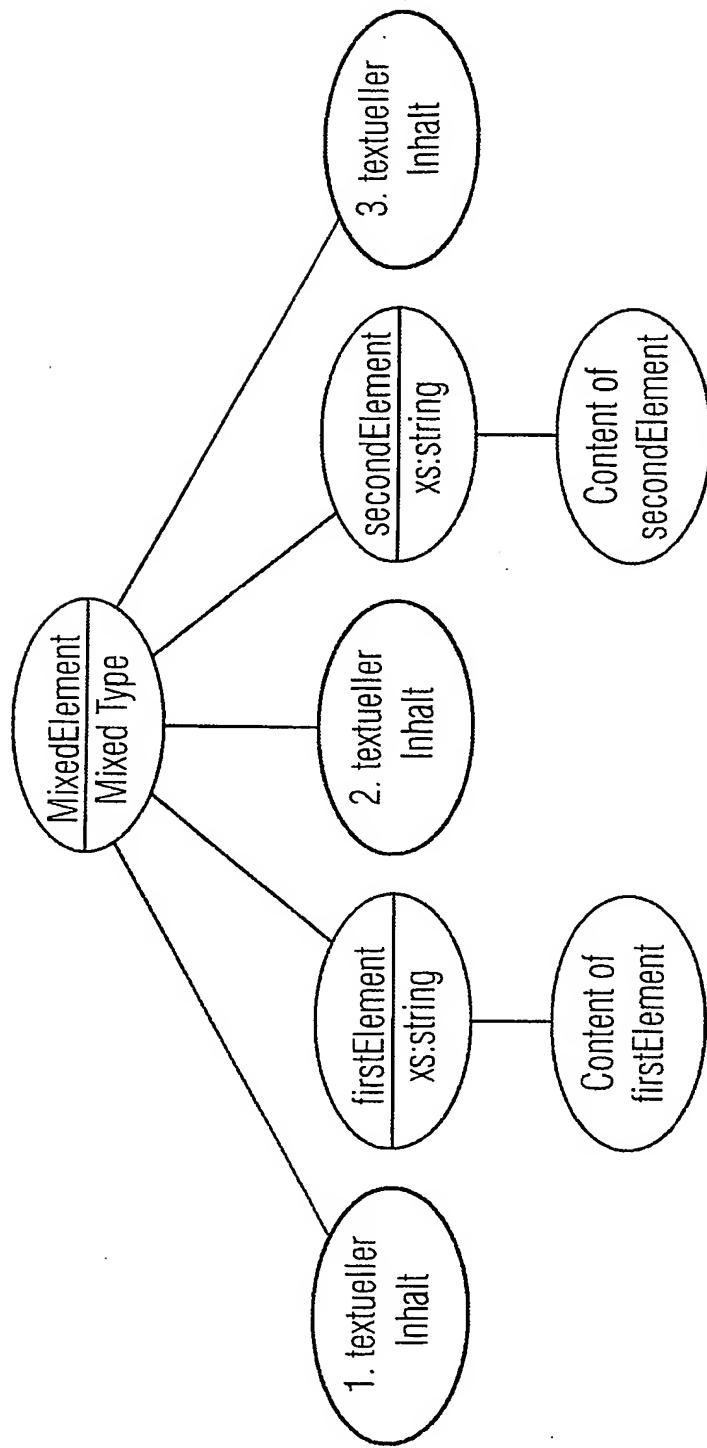
FIG 3

```
<?xml version="1.0" encoding="UTF-8"?>
<Example xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="C:\data\Awaj\BiM\mixed\mixed.xsd">
<MixedElement>
  1. textueller Inhalt
  <firstElement>Content of firstElement</firstElement>
  2. textueller Inhalt
  <secondElement>Content of secondElement</secondElement>
  3. textueller Inhalt
</MixedElement>
</Example>
```

THIS PAGE BLANK (USPTO)

3/4

FIG 4



THIS PAGE BLANK (USPTO)

4/4

FIG 5

Struktur Code	Adressierter Knoten
00	Vaterknoten
01	textueller Inhalt
10	firstElement
11	secondElement

FIG 6

Position Code	Adressierter Knoten
-	Vaterknoten
00, 01, 10	textueller Inhalt
-	firstElement
-	secondElement

THIS PAGE BLANK (USPTO)